Foundation Course on Statistics and Optimization for Data Scientists

About the Course:

To provide adequate in-depth knowledge on the following topics so as to make the participants to confidently handle practical problems through modelling and formulation, and make them confident in learning advanced topics such as multivariate analysis, Bayesian data analysis, machine learning etc. which are useful for Big Data Analytics on their own.

Mode of Instruction:

Weekly two sessions – each of 3 hours duration, three days gap between every two consecutive sessions so that participants work on the home assignments during the gaps. All topics will be covered with live practical examples so that the real use of statistical applications is realized while learning the basic theory.

Qualification:

- ✓ Basic Mathematics at intermediate level.

Duration of the course:

- ✓ One year

Mode of course delivery

- ✓ Class Room(Venue:Lingampally)/Online Training

Faculty Details:

- ✓ A team of faculty having an average 20 + years' experience in the data analysis across various industries and training.

Subjects and Topics

1. **Matrix Algebra (30 Hours)**
    a. **Vector Spaces**: Arithmetic, bases, dimension, system of equations, inner product.
    b. **Matrices**: Basic operations, row and column spaces, null space, rank, trace, determinant, diagonal reductions, orthogonal matrices, transformations, inverse and g-inverse, idempotent matrices, partitioning of matrices and manipulations.
    c. **Quadratic Forms**: Classification of QF (positive semidefinite, positive definite, negative and negative semidefinite), transformation to diagonal forms, eigen values and vectors, and spectral decomposition.
    d. **Projection Operator**: Projection of a vector onto a subspace and its application in regression analysis.

2. **Probability (40 Hours)**
    a. **Uncertainty and its measurement:** Examples of random phenomena
    b. **Probability – Basic Definitions:** Outcomes, events, disjoint events and additive property
    c. **Conditional Probability:** Definition and applications, Bayes theorem, and Bayesian approach.
    d. **Independence:** Definitions and some examples.
    e. **Combinatorics:** Ordered samples, permutations, combinations, partitions, union of events, matching problems, occupancy problems, number of empty boxes.
    f. **Random Variables:** Random variables and vectors, density, distribution function, independence, standard discrete distributions (Bernoulli, binomial, geometric, negative binomial, multinomial, Poisson), sums of independent random variables.
    g. **Expectations:** Definition of expectation and its properties, moments, variance of sum of independent random variables, covariance and correlation, Chebyshev's inequality.
    h. **Continuous Random Variables**: Definition, density and distribution functions of continuous random variables, transformations.
    i. **Standard Continuous Distributions:** Uniform, normal, exponential, gamma, Chisquare, students t, and F.
    j. **Joint Distributions**: Distributions of sums of random variables, conditional densities, posterior distribution.
    k. **Multivariate Normal Distribution**: Bivariate and multivariate normal distributions, conditional distributions of multivariate normal.
    l. **Expectations**: Central Limit Theorem, law of large numbers, moment generating functions.

3. **Stochastic Processes (30 Hours)**
    a. **Markov Chains**: Definition and examples of Markov Chains, stationarity, initial and stationary distributions, classification of states, steady state probabilities and applications of Markov chains.
    b. **Continuous Time Markov Chains:** Pure jump processes, birth and death chains, Poisson process, second order processes, mean and covariance functions, auto correlation function, Gaussian process and White Noise.

4. **Statistical Inference (30 Hours):**
    a. **Basic Statistics:** Types of problems and principles, risk and loss functions.
    b. **Estimation:** Unbiased and efficient estimators, sufficiency and complete estimators, maximum likelihood estimators, Bayesian estimation, confidence and prediction intervals.
    c. **Tests of Hypotheses:** Null and alternate hypothesis, critical region and test statistics, power of a test, standard tests (comparison of proportions, comparison of variances, z-test, 1-sample and two sample t-tests, paired t-test).
    d. **Analysis of Variance:** One and two-way classifications.
    e. **Regression analysis**: Linear models and regression analysis.

5. **Optimization (30 Hours):**
    a. **Optimization Problems**: Examples of optimization problems and mathematical models.
    b. **Linear Programming**: Principles, examples and solutions using software.
    c. **Standard Models**: Transportation and assignment models, Network models, travelling salesman problem.
    d. **Some Live Applications**: Production and inventory problems, human resource management problems, material optimization problems.

6. **Time Series and Forecasting (30 Hours):**
    a. **Time series data**: Some simple forecasting methods, Time series graphics, Seasonal or cyclic trends, Autocorrelation, Forecast residuals, White noise, Evaluating forecast accuracy, Exponential smoothing
    b. **ARIMA Models**: Common ARIMA models, ACF and PACF of seasonal ARIMA models, Transformations, Stationarity, Ordinary differencing, Seasonal differencing,
    c. **State space models**: Simple structural models, Linear Gaussian state space models, Kalman filter, ARIMA models in state space form, Kalman smoothing.

7. **Reliability and Survival Analysis (30 Hours):**
    a. **Reliability**: Simple reliability system (parallel and series configurations, standby systems, k-out-of-n systems, cold and hot standbys), Time dependent reliability models (hazard and reliability functions, mean time to failure and mean time between failures), Stress-strength reliability models.
    b. **Availability and Maintainability**: Analysis of reparable, irreparable and periodically maintained systems, Contributions to unavailability.
    c. **Survival Analysis**: Life testing, types of censoring, parametric distribution analysis.